QUANTITATIVE STUDY OF UNDERSAMPLED RECOVERABILITY FOR SPARSE IMAGES IN COMPUTED TOMOGRAPHY

J. S. JØRGENSEN[*], E. Y. SIDKY[†], P. C. HANSEN[*] , AND X. PAN[†]

Abstract. Image reconstruction methods based on exploiting image sparsity, motivated by compressed sensing (CS), allow reconstruction from a significantly reduced number of projections in X-ray computed tomography (CT). However, CS provides neither theoretical guarantees of accurate CT reconstruction, nor any relation between sparsity and a sufficient number of measurements for recovery. In this paper, we demonstrate empirically through computer simulations that minimization of the image 1-norm allows for recovery of sparse images from fewer measurements than unknown pixels, without relying on artificial random sampling patterns. We establish quantitatively an average-case relation between image sparsity and sufficient number of measurements for recovery, and we show that the transition from non-recovery to recovery is sharp within well-defined classes of simple and semi-realistic test images. The specific behavior depends on the type of image, but the same quantitative relation holds independently of image size.

Key words. Computed Tomography, Image Reconstruction, Sparse Approximation, Compressed Sensing, Recoverability, Inverse Problems

AMS subject classifications. 90C90, 15A29, 94A08, 44A12

1. Introduction. In X-ray computed tomography (CT) an object is to be reconstructed from projections obtained by measuring the attenuation of X-rays sent through the object. The forward problem, which describes how the projection data arises from the object, can be modelled either with an analytical formulation or an algebraic formulation. In the analytical formulation, the image is reconstructed by explicitly evaluating closed-form analytical inverses of the forward operator. Examples of analytical reconstruction methods are filtered back-projection [21] and the Feldkamp-Davis-Kress method for cone-beam CT [14]. Their main advantages are low memory demands and computational efficiency, which make them the current methods of choice in commercial CT scanners [3, 22].

In the algebraic formulation the forward operator is fully discretized, which leads to a large system of linear equations that must be solved in a stable fashion, typically by means of an iterative algorithm. The main advantage of the algebraic formulation is flexibility: It can handle geometries for which no analytical inverse is available, such as non-standard scanning geometries, and it allows for incorporation of various forms of prior knowledge about the object. Within the algebraic formulation there are many different kinds of reconstruction methods; some, such as the algebraic reconstruction technique (ART) and variants compute an approximate solution to the discretized linear system. These methods allow for simple constraints such as non-negativity. In optimization-based or variational methods the image is reconstructed by extremizing an objective function chosen to measure image desirability. These methods allow for incorporation of many types of prior knowledge into the objective function or constraints, such as image smoothness and knowledge of the statistical noise model.

Motivated by the desire to reduce the exposure to radiation, there is a growing interest in low-dose CT, cf. [31] and references therein. This is relevant in medical imaging to reduce the risk of radiation-induced cancer, and in biomedical imaging

[*]Department of Informatics and Mathematical Modeling, Technical University of Denmark, Building 321, Asmussens Allé, 2800 Kgs. Lyngby, Denmark ({`jakj,pch`}`@imm.dtu.dk`).

[†]Department of Radiology, University of Chicago, 5841 S. Maryland Ave., Chicago, IL 60637, USA ({`sidky,xpan`}`@uchicago.edu`).

where high doses can damage the specimen under study. It is recognized that the data reduction arising from low-dose imaging can sometimes be compensated for by incorporating prior information about the image, which therefore calls for the algebraic formulation, typically in the form of optimization-based reconstruction.

Developments in compressed sensing (CS) [6, 10] show potential for a reduction in data while maintaining or even improving reconstruction quality. This is made possible by exploiting image sparsity; loosely speaking, if the image is "sparse enough", i.e., having "sufficiently few" nonzero elements in some representation, it admits accurate reconstruction from "undersampled" data. We will refer to such methods as *sparsity-exploiting methods*.

A typical cross-section image of the human body consists of well-separated areas of relatively homogeneous tissue, and same is true for poly-crystalline materials (such as metals) and for many other applications. Such images can be seen as having an approximately sparse gradient, and this property can be exploited in a reconstruction algorithm by minimizing the total variation (TV) semi-norm [28]. Studies using simulated as well as clinical data have demonstrated that sparsity-exploiting methods indeed allow for reconstruction from fewer projections [4, 8, 16, 27, 29, 30]. In spite of these positive results, we still lack a fundamental understanding of conditions under which such methods can be expected to perform well in CT.

The present paper investigates a possible relation between the image sparsity and the number of CT views sufficient for recovery of the underlying true image. To simplify our discussion and analysis we focus on reconstruction of sparse images, i.e., images with a large number of zero pixels. We can apply ℓ_1-norm minimization of the image to enforce this kind of sparsity. These studies are interesting in their own right (see [19] for an example of ℓ_1-norm based CT reconstruction of blood vessels), and they set the stage for forthcoming studies of TV and other types of sparsity.

We are unaware of theoretical results that cover the mathematical model for CT. Inspired by work of Donoho and Tanner [9] we therefore use computer simulations to study *recoverability* by ℓ_1-norm minimization within well-defined classes of sparse images, that is, the average number of projections sufficient for recovering an image as function of the image sparsity. To this end, we introduce for a single image the concepts of *strong recovery* meaning that the image is reconstructed accurately (to within a chosen numerical accuracy), and *weak recovery* meaning that only an image with same ℓ_1-norm as the original is found. We also introduce the *relative sparsity-sampling* diagram for studying these quantities as function of image sparsity and number of projections, and we argue that these tools are of general value for studies of sparsity-exploiting methods. An advantage of our empirical approach (instead of relying on specific theoretical results) is that we can use the same technique to characterize systems of increasing levels of realism. Our findings for X-ray CT can be summarized as follows:

1. Image recovery is possible with a structured sampling pattern as used in commercial CT scanners.
2. There is a quantitative relation between image sparsity and sufficient sampling for recovery.
3. There is a sharp transition from non-recovery to recovery.
4. The relation varies with respect to different image classes.
5. The relation holds independently of the image dimension.

Another interesting result is that the gain in using a sparsity-exploiting method is significant, even for images with relatively many nonzeros.

Our paper is organized as follows. Section 2 gives the problem formulation and the sparsity-exploiting reconstruction method, and it introduces the concept of recoverability in CT. Section 3 describes our numerical simulations, including details of the CT imaging model, generation of sparse phantoms, and how to robustly solve the reconstruction optimization problems. Section 4 presents an overview of our results, and we conclude with a discussion of the results in Section 5.

2. Sparsity-exploiting reconstruction methods for CT. The purpose of this brief section is to define the notation, the algebraic formulation, and the reconstruction method used throughout the study.

2.1. Constrained algebraic reconstruction. We consider the discrete inverse problem of recovering a signal $x_{\text{orig}} \in \mathbb{R}^N$ from (usually noisy) measurements $b \in \mathbb{R}^M$. The imaging model, which is assumed to be linear and discrete-to-discrete [2], relates the image and the data through a *system matrix* $A \in \mathbb{R}^{M \times N}$,

$$b = A\,x + e, \tag{2.1}$$

where $e \in \mathbb{R}^M$ represents additive noise, and the elements of $x \in \mathbb{R}^N$ are pixel values stacked into a vector. To solve (2.1) it is often necessary to impose some sort of regularization in order to reduce noise amplification in the inversion as well as to restrict the set of solutions in case of an underdetermined and/or rank deficient problem. A very common type of regularization takes the form: $\min_x\ J(x)$ s.t. $\|A\,x - b\|_2^2 \leq \epsilon^2$, where $J(x)$ is a *regularizer*, i.e., a function selected to impose some condition on the image that reflects prior knowledge or assumptions. In this work we use $J(x) = \|x\|_1$, which is known in many cases to produce a sparse x, as discussed below.

The regularization parameter ϵ reflects the amount of noise in the data, and in the limit $\epsilon \to 0$ we obtain the equality-constrained problem:

$$\textbf{L1:} \qquad \min_x\ \|x\|_1 \qquad \text{s.t.} \qquad A\,x = b. \tag{2.2}$$

The corresponding problem **L2** with $J(x) = \|x\|_2^2$, which corresponds to Tikhonov regularization, gives the unique minimum-norm solution, i.e., the vector of minimal ℓ_2-norm among all vectors satisfying $A\,x = b$. The problem **L1** can have non-unique solutions, an issue which we will address in Section 3.4.

The inequality-constrained problem is of more practical interest than (2.2) because it allows for noisy and inconsistent measurements, but its solution depends in a complex way on the noise and inconsistencies in the data, as well as the choice of the parameter ϵ. Studies of the equality-constrained problem (2.2), on the other hand, provide an basic understanding of a given regularizer's reconstruction potential, independent of specific noise. Therefore, we focus in the present study on the ideal equality-constrained problem. This means that we do not consider noisy data, uncertainties in the system matrix, or the influence of a regularization parameter.

2.2. Recoverability of problem instances. The interest in **L1** (as well as TV and other sparsity-exploiting methods) is motivated by recent developments in CS demonstrating that it is possible to recover x_{orig} from a reduced number of measurements [6, 10]. The underlying assumption is that the image is has few non-zero pixels, or that it is sparse in some other representation (such as after applying a discrete gradient transformation to the image). We call a vector $x \in \mathbb{R}^N$ with k non-zero

elements k-sparse and we define

$$\textbf{relative sparsity:} \qquad \kappa = k/N. \tag{2.3}$$

Moreover, we call a tuple $(x_{\mathrm{orig}}, A, b = A\,x_{\mathrm{orig}})$ a *problem instance* and say that it is *recoverable* if the solution $x^\star$ to **L1** is identical to x_{orig}.

There has been much work on establishing results stating for which matrices A problem **L1** is capable of recovering the sparsest solution, cf. [5] and references therein. We give an example of such a theorem based on the mutual coherence μ_A of a matrix A: For a full-rank $A \in \mathbb{R}^{M \times N}$ with $M < N$, if a k-sparse solution x to $A\,x = b$ satisfies

$$k < \frac{1}{2}\left(1 + \frac{1}{\mu_A}\right), \qquad \mu_A = \max_{1 \le i,j \le N} \frac{|a_i^T a_j|}{\|a_i\|_2 \cdot \|a_j\|_2}$$

(where a_ℓ is the ℓth column of A), then it is the unique solution to **L1**; a smaller μ_A leads to a larger bound on the sparsity of a signal that is guaranteed to be recovered. Many theoretical recovery results exist, most notably relying on the so-called "spark" [11] and restricted isometry property [7] of a matrix. While some results are deterministic, many of them are probabilistic in the sense that if the elements of A are selected at random from certain probability distributions, then with "overwhelming probability" **L1** will recover the original signal [6].

It is generally accepted that these theoretical results are of limited practical use [13], since the requirements are generally NP-hard to check [24], and/or they provide very pessimistic bounds on the sparsity of signals that can be recovered. Better results are available for certain special matrices, such as those with elements drawn from a Gaussian distribution, but those results do not carry over to the system matrices encountered in CT. Instead, recoverability can be studied empirically. Such studies have been conducted for certain practical systems, (see, e.g., [1, 23]), but we are unaware of any systematic recoverability studies specifically for CT system matrices.

Our empirical study is inspired by the work by Donoho and Tanner (DT) [9] who studied empirical recoverability using a *phase diagram* of the (δ, ρ)-plane, where:

$$\textbf{undersampling fraction:} \quad \delta = M/N, \qquad \textbf{sparsity fraction:} \quad \rho = k/M. \tag{2.4}$$

For certain classes of random matrices DT were able to prove existence of a sharp transition from non-recovery to recovery, and verify the result in empirical studies. Although we do not derive similar theoretical results for CT matrices, we can still conduct similar empirical recoverability studies using the DT phase diagram and the related relative sparsity-sampling (RSS) diagram, which we introduce in Section 4.

We note that it is possible to construct examples of k-sparse vectors for small k that cannot be recovered from CT measurements [12, 25], implying that we cannot hope to obtain useful results on guaranteed recovery of *all* k-sparse images. However, these constructed vectors might be pathological and very different from actual images occurring in practice, and for this reason we are more interested in determining *average-case* recovery results for specific classes of images.

We will empirically establish a quantitative relation between the number of measurements and the sparsity of x_{orig} sufficient for recoverability. In order to do that, we conduct randomized simulations where we generate ensembles of images of varying sparsity and CT system matrices corresponding to varying number of projections; then we use **L1** for reconstruction, and we check for recovery. Since different phantoms of same class and sparsity might require different number of views to be recovered, we are interested in the average-case recovery over the phantom ensembles.

3. Experiment design. In this section we describe the test problems used in our studies, as well as our approach to solving the reconstruction problem numerically.

3.1. CT imaging geometry. Different scanning geometries are used depending on the manufacturer and scanner type. When designing a reconstruction algorithm using a given set of measured projections, one can adjust the number of pixels to trade-off resolution for signal-to-noise ratio, re-bin or interpolate the data to obtain additional "data", use other basis functions than pixels, etc.

We consider a generic geometry to serve as an example: the standard 2D fan-beam geometry with equi-angular views as illustrated in Figure 3.1. Due to rotational symmetry we restrict the image to be within a circular mask inside an $N_{\text{side}} \times N_{\text{side}}$ square domain; the number of pixels in the masked image is approximately $N = \lceil \pi/4 \cdot N_{\text{side}}^2 \rceil$. The source-to-detector distance is $2N_{\text{side}}$, and the fan-angle $28.07°$ is set to precisely illuminate the circular mask. The number of views (or projections) is denoted N_{v}. The rotating detector is assumed to consist of $N_{\text{b}} = 2N_{\text{side}}$ bins, so the total number of measurements is $M = 2N_{\text{side}}N_{\text{v}}$. Each measurement is modeled with a linear equation of the form

$$b_i = \sum_{j=1}^{N} A_{ij}\, x_j, \qquad i = 1, \ldots, M, \tag{3.1}$$

where A_{ij} is the path length of the ith ray through the jth pixel (this is the so-called line intersection method of discretizing the forward operator). The system matrix A is then $M \times N$, and it is computed by means of the MATLAB package AIR Tools [17].
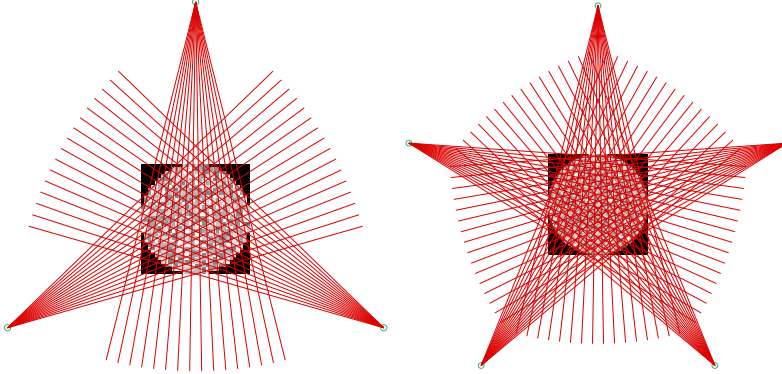

FIG. 3.1. *The fan-beam geometry with views from full angular range. Pixels outside the disk (shown as black) are not considered part of the image. Left: disk diameter $N_{\text{side}} = 32$ and $N_v = 3$ views. Right: disk diameter $N_{\text{side}} = 64$ and $N_v = 5$ views. To improve the visualization, the number of rays per view has been reduced compared to the choice $M = 2N_{\text{side}}$ used in the simulations.*

3.2. Sparse phantom classes and generation of instances. By a *class of phantoms* we mean a set of phantom images that are described by a set of specifications, in such a way that we can generate random realizations from the class. We refer to such an image as a *phantom instance* from the class, and multiple phantom instances from the same class are said to form a *phantom ensemble*.

3.2.1. The "spikes" and "signedspikes" classes. For the **spikes** class, given an image size N and a target relative sparsity κ, we generate a phantom instance
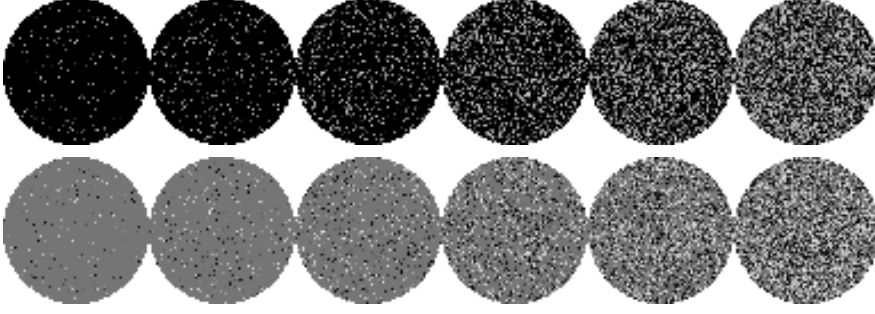
FIG. 3.2. *Sparse phantom image instances of class* **spikes** *(1st row, black is 0, white is 1), and of class* **signedspikes** *(2nd row, black is −1, white is 1). Relative sparsity from left to right is* $\kappa = 0.05, 0.10, 0.20, 0.40, 0.60, 0.80$.

$x \in \mathbb{R}^N$ as follows: Given a zero vector x, randomly select $k = \text{round}(\kappa N)$ pixel indices, and set each selected pixel value in x to a random number from a uniform distribution over $[0, 1]$. Figure 3.2 shows examples of **spikes** phantom instances for varying κ. This class is deliberately designed to be as simple as possible and it does not mimic any particular application. It is used to study the generic case of having a sparse image, and thus provide basic understanding before pursuing any specific application with more realistically looking phantoms.

The **signedspikes** class is essentially identical to the **spikes** class, except the nonzero pixel values are uniformly distributed over $[−1, 1]$. In standard CT reconstruction the attenuation coefficient is always non-negative, but in certain applications a background attenuation is subtracted, thereby causing x to have both positive and negative values. As we will show in Section 4, the modification to allow negative pixel values leads to a significant change in the recoverability.

3.2.2. The "p-power" class. As a phantom class that models a more realistic type of images, we choose a model of breast tissue. The general idea is to introduce some structure to the non-zero pixels by creating correlation between neighboring pixels. Our procedure is based on [26] followed by thresholding to obtain many pixels that are strictly equal to 0. The amount of structure is governed by a parameter p and we refer to this parameterized family of phantom classes by p**-power**:

1. Create a $N_{\text{side}} \times N_{\text{side}}$ phase image P with values drawn from a Gaussian distribution with zero mean and unit standard deviation.
2. Create an $N_{\text{side}} \times N_{\text{side}}$ amplitude image U with pixels

$$U(i, j) = (u_i^2 + u_j^2)^{-p/2}, \quad i, j = 1, \ldots, N_{\text{side}},$$

where $u_\ell = -1 + (2\ell - 1)/N$, $\ell = 1, \ldots, N_{\text{side}}$.
3. For all pixels (i, j) compute $F(i, j) = U(i, j)e^{2\pi i P(i,j)}$.
4. Compute the magnitude of the 2D inverse discrete Fourier transform of F.
5. Restrict this square image to the disk-shaped mask.
6. Determine the number of non-zeros $k = \text{round}(\kappa N)$.
7. Keep the k largest pixel values and set the rest to 0.

Figure 3.3 shows examples of phantoms from classes **1.0-power** and **2.0-power**. We see that the **1.0-power** phantoms have more structure than the **spikes** phantoms, and **2.0-power** has even more structure.
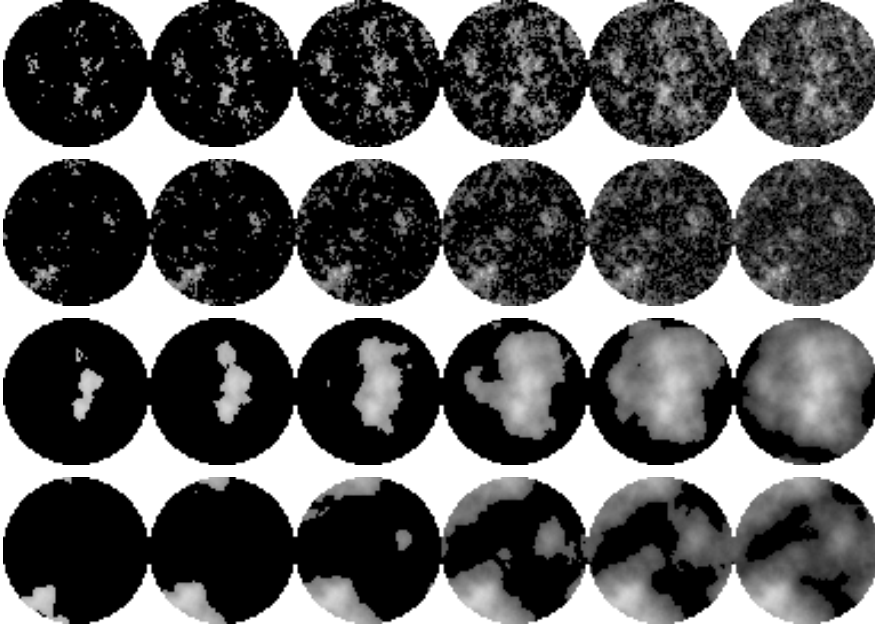
FIG. 3.3. *Two sets of phantom image instances of class* **1.0-power** *(1st and 2nd rows, black is 0), and of class* **2.0-power** *(3rd and 4th rows, black is 0). Relative sparsities from left to right are $\kappa = 0.05, 0.10, 0.20, 0.40, 0.60, 0.80$.*

3.3. Robust software for solving the optimization problems. The robustness of a numerical solution for deciding recovery depends on the solution accuracy. False conclusions may result from incorrect or inaccurate numerical solutions. To robustly solve the optimization problems **L1** and **L2** we must therefore use a numerical method which gives a clear indication of whether a correct solution, within a given accuracy, has been computed. Our choice is the package MOSEK [20], which uses a primal-dual interior-point method suitable for large-scale sparse problems. MOSEK is equipped with numerous sophisticated features to handle numerical difficulties, and it issues warnings when it fails to compute an accurate solution. In all problem instances considered in our simulation studies, MOSEK managed to return a certified accurate solution.

Solving **L2** with MOSEK is straightforward, since rewriting $\|x\|_2^2 = x^T x$ allows us to cast **L2** as a quadratic program, which is readily solved by MOSEK. To solve **L1** using MOSEK we cast it as a linear program (LP) by introducing a new variable vector $w \in \mathbb{R}^N$ bounding the magnitude of x's elements from above and below, $-w \leq x \leq w$, which implicitly makes w non-negative. Minimizing $\mathbf{1}^T w$ for $\mathbf{1} \in \mathbb{R}^N$, while respecting the original constraint $A x = b$ as well as $-w \leq x \leq w$, is an LP which is reliably solved by MOSEK.

3.4. Strong and weak L1 recovery. While **L2** has a unique solution because $\|x\|_2^2$ is strictly convex, it is not necessarily the case for **L1**. In a specific imaging application, the system matrix and the image are given, and the arising problem instance might cause L1 to not have a unique solution. To handle this issue, given a specific problem instance $(A, x_{\mathrm{orig}}, b = A x_{\mathrm{orig}})$ we consider two different notions of recovery for the computed solution $x^\star$ to **L1**:

- **Strong recovery**: $x^\star$ is identical to the original: $x^\star = x_{\mathrm{orig}}$.
- **Weak recovery**: $x^\star$ has the same optimal value as x_{orig}: $\|x^\star\|_1 = \|x_{\mathrm{orig}}\|_1$.

To motivate the distinction between different types of recovery we consider the possible outcomes from solving **L1**:

1. $\|x^\star\|_1 > \|x_{\mathrm{orig}}\|_1$. This is not possible, since the smaller objective of x_{orig} would cause the robust optimization algorithm to return x_{orig} in favor of $x^\star$.
2. $\|x^\star\|_1 < \|x_{\mathrm{orig}}\|_1$. Another image, $x^\star$, with smaller objective than x_{orig} exists, so there is no hope for recovery with **L1**.
3. $\|x^\star\|_1 = \|x_{\mathrm{orig}}\|_1$. This is weak recovery, so we know that x_{orig} is at least *a* minimizer. We do not know whether the found minimizer $x^\star$ is identical to x_{orig}, because we do not know whether the **L1**-solution is unique. If x_{orig} is indeed the unique **L1**-solution, then x_{orig} is sometimes called *identifiable* [12]. It is possible to test for **L1**-identifiability by verifying a set of necessary and sufficient conditions [15], but for general $J(x)$ we are not aware of any general test, and we leave uniqueness tests for future work. Given weak recovery we test for strong recovery. The possible outcomes are:

   - $x^\star = x_{\mathrm{orig}}$. With strong recovery, the original is recovered. This does not imply identifiability, so we can not be certain that another choice of optimization algorithm would also yield recovery.
   - $x^\star \neq x_{\mathrm{orig}}$. Without strong recovery, a different image with same objective as the original was found, so the **L1** solution can not be unique. This, in itself, is valuable information. Additionally, it gives two strategies for potentially recovering the original. First, a different optimization algorithm may be able to recover the image (for instance, an interior-point method tends to select solutions in the interior of faces on the LP simplex, while a simplex-method selects vertices of the LP simplex.) Second, it may be possible to restrict the feasible set by introduction of constraints, such as nonnegativity, reflecting any prior knowledge of the image, and thereby recover the original.

We want to emphasize here that the decision of strong recovery is dependent on the choice of optimization algorithm and its internal parameters (such as the initial point and stopping rules). In contrast, weak recovery is *independent of the algorithm*, as long as the optimization algorithm is robust. In our simulations, we will therefore — in addition to strong recovery results — also report weak recovery results.

3.5. Simulations. Given the imaging model, the procedure for generating sparse phantom classes, and a robust optimization algorithm, we are in the position to carry out randomized simulation studies of recoverability within a phantom class, as a function of relative sparsity and number of views. The design of a single basic simulation consists of the following steps: 1) Generate an instance $(A, x_{\mathrm{orig}}, b = Ax_{\mathrm{orig}})$, 2) solve **L1** numerically to obtain $x^\star$, and 3) test for strong recovery ($x^\star = x_{\mathrm{orig}}$) and weak recovery ($\|x^\star\|_1 = \|x_{\mathrm{orig}}\|_1$) using

$$\frac{\|x^\star - x_{\mathrm{orig}}\|_2}{\|x_{\mathrm{orig}}\|_2} < \epsilon \qquad \text{and} \qquad \frac{\big|\|x^\star\|_1 - \|x_{\mathrm{orig}}\|_1\big|}{\|x_{\mathrm{orig}}\|_1} < \epsilon, \tag{3.2}$$

where the threshold ϵ is chosen based on the accuracy of the optimization algorithm. We used $\epsilon = 10^{-4}$, and this choice is studied in Section 4.4.

4. Simulation results. We start by introducing some notation that is useful for the following discussion. For a given problem instance, the *sufficient view number*

$N_\mathrm{v}^\mathrm{suf}$ denotes the smallest number of views that gives a full-rank matrix A and thus recovery with **L2**. The **L1** *recovery view number* $N_\mathrm{v}^\mathbf{L1}$ denotes the smallest number of views for which strong recovery is observed for all $N_\mathrm{v} \geq N_\mathrm{v}^\mathbf{L1}$. Using $N_\mathrm{v}^\mathrm{suf}$ as a reference point for full sampling, we define two useful quantities:

$$\textbf{relative sampling:} \quad \mu \quad = N_\mathrm{v}/N_\mathrm{v}^\mathrm{suf}, \tag{4.1}$$

$$\textbf{relative sampling for recovery:} \quad \mu^\mathbf{L1} = N_\mathrm{v}^\mathbf{L1}/N_\mathrm{v}^\mathrm{suf}, \tag{4.2}$$

cf. the relative sparsity κ defined in (2.3).

One of our main contributions is the so-called relative sparsity-sampling (RSS) diagram, introduced in Section 4.2, in which the recovery percentage over an ensemble of phantoms is plotted as function of the relative sparsity and the relative sampling. Strong and weak recovery RSS diagrams reveal a sharp transition from non-recovery to recovery and a monotonically increasing relation between relative sparsity and relative sampling for recovery. The subsequent sections study how the RSS diagrams change with variations to the recovery criteria, image size, phantom class and diagram resolution; a summary of results is given in Section 4.8.
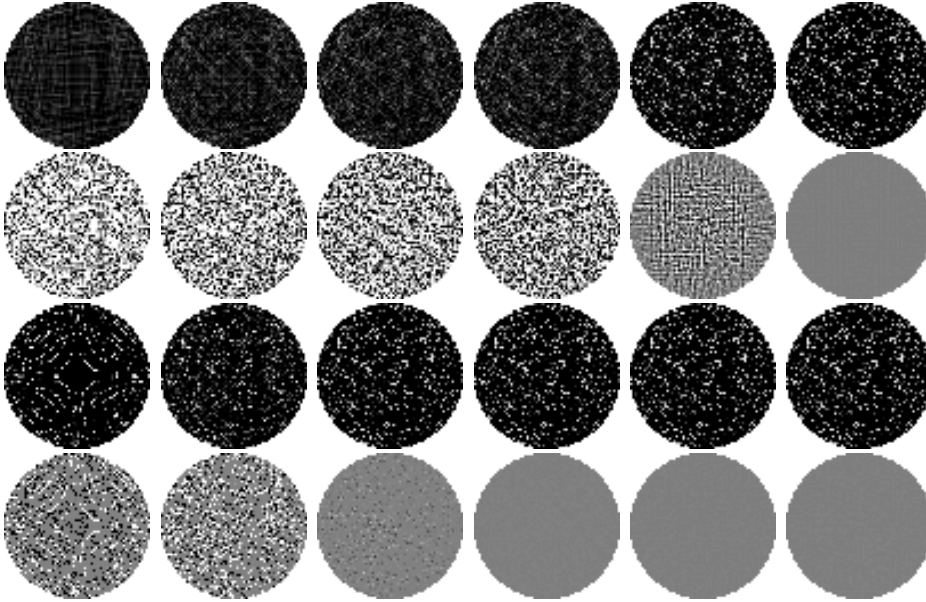

FIG. 4.1. *Reconstructions of the* **spikes** *phantom with* $N_\mathrm{side} = 64$, *relative sparsity* $\kappa = 0.20$. *1st row:* **L2** *reconstructions (black is* 0, *white is* 1*). 2nd row:* **L2** *minus original image (black is* -0.1, *white is* 0.1*). 3rd row:* **L1** *reconstructions (black is* 0, *white is* 1*). 4th row:* **L1** *minus original image (black is* -0.1, *white is* 0.1*). Columns:* 4, 8, 10, 12, 24 *and* 26 *views.*

4.1. Recovery from undersampled data. We first establish that **L1** is capable of recovering an image from undersampled CT measurements, and we compare with the **L2** reconstruction. We use a phantom x_orig from the **spikes** class with $N_\mathrm{side} = 64$, leading to $N = 3228$ pixels inside the mask. The relative sparsity is set to $\kappa = 0.20$, which yields 646 non-zeros. We consider reconstruction from data corresponding to $4, 6, 8, \ldots, 32$ views; the smallest and largest system matrices are of sizes 512×3228 and 4096×3228, respectively. Selected reconstructed images from
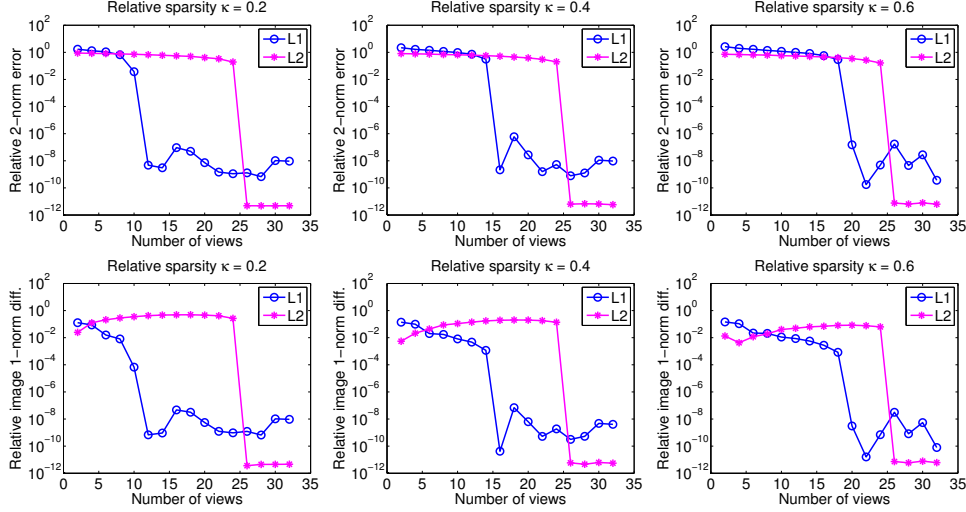
FIG. 4.2. *Strong (top row) and weak (bottom row) recovery criterion measures from (3.2) vs. view numbers for* **L1** *and* **L2** *reconstruction of* **spikes** *phantoms with relative sparsity values $\kappa = 0.2$, 0.4 and 0.6.*

solving **L1** and **L2** are shown in Figure 4.1 along with the difference image w.r.t. x_{orig} to better visualize the transition to recovery.

Recall that the minimum-norm **L2** solution is typically non-sparse [13]. Therefore, we expect to need a full-rank system matrix for **L2** to recover the original sparse image, and our results confirm this: the **L2** reconstructions gradually improve with more views but recovery is not observed until $N_{\text{v}} = N_{\text{v}}^{\text{suf}} = 26$. At $N_{\text{v}} = 24$, the size is 3072×3228 and $\text{rank}(A) = 3052$, and the minimum-norm **L2** solution is not the original. At $N_{\text{v}} = 26$, the matrix is 3328×3228 and full-rank, so x_{orig} is recovered.

For **L1**, recovery occurs already at $N_{\text{v}} = N_{\text{v}}^{\textbf{L1}} = 12$, where A has size 1536×3228 and rank 1524. Evidently, in spite of the large null space of A, in this case, **L1** selects the original image. When N_{v} increases, **L1** continues to recover the original up to and beyond $N_{\text{v}} = N_{\text{v}}^{\text{suf}} = 26$, where the matrix becomes full-rank. Figure 4.1 thus demonstrates the potential of **L1** recovery from undersampled CT measurements.

In order to investigate quantitatively a possible relation between image sparsity and sufficient sampling for **L1** recovery we repeat the $\kappa = 0.2$ experiment for $\kappa = 0.4$ and 0.6. Figure 4.2 shows the strong and weak recovery criterion measures from (3.2) for **L1** and **L2** as a function of N_{v}. For **L2** the behavior is independent of the relative sparsity, as expected. For **L1**, on the other hand, $N_{\text{v}}^{\textbf{L1}}$ takes the values 12, 16 and 20, indicating a very simple relation between sparsity and **L1** recovery view number. For the problem instances considered here strong and weak recovery coincide but, as we will see below, that is not always the case.

4.2. Recoverability studies using the DT diagram. In general, we can not expect all phantom instances of same relative sparsity to have the same $N_{\text{v}}^{\textbf{L1}}$, and in fact we observe some variation. One way to study this variation is through the DT phase diagram described in Section 2.2. For the **spikes** class with $N_{\text{side}} = 64$ we consider reconstruction with undersampling fraction $\delta = M/N$ for $M = N_{\text{v}} N_{\text{b}}$, where $N_{\text{b}} = 2N_{\text{side}}$ and $N_{\text{v}} = 2, 4, 6, \ldots, 32$. At each N_{v} we consider sparsity fractions $\rho = k/M = 1/16, 2/16, \ldots, 16/16$. We test for strong and weak recovery for reconstruction
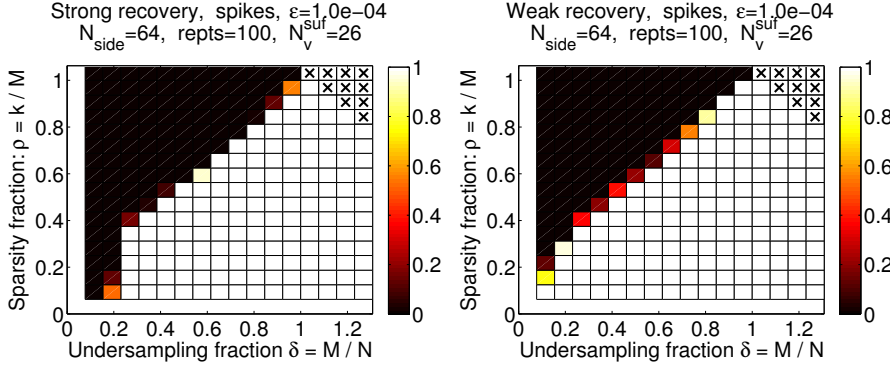
Fig. 4.3. *DT phase diagrams for strong and weak recovery of* **spikes** *phantoms,* $N_{\mathrm{side}} = 64$.

using the same system matrix A for 100 phantom instances at each (δ, ρ).

The resulting diagrams with the percentage of instances recovered at each (δ, ρ) are shown in Figure 4.3. The strong recovery DT phase diagram shows that at very low sampling ($N_{\mathrm{v}} = 2$ and 4, i.e., $\delta = 0.08$ and 0.16) no phantoms can be recovered. An important observation is the very sharp phase transition from non-recovery to recovery, meaning that the variation of $N_{\mathrm{v}}^{\mathbf{L1}}$ within the same phantom class is very limited. To the best of our knowledge, a similar analysis has not been done for CT-matrices before, and we therefore believe the sharp transition to be a new observation. Interestingly, in the weak recovery DT phase diagram, the transition extends all the way to the low sampling cases; showing potential for recovery from very few samples, provided one can find a way to restrict the solution to be unique.

Normalizing the sparsity by the number of measurements causes a small problem for the diagram. When $\delta > 1$, values of ρ close to 1 can lead to $k > N$, for which it is impossible to construct instances. Those cases are shown with $\times$-symbols in the DT diagrams.

4.3. The relative sparsity-sampling (RSS) diagram. As an alternative to the DT phase diagram we can visualize the percentage of instances recovered in the (κ, μ)-plane for the relative sparsity $\kappa = k/N$ and the relative sampling $\mu = N_{\mathrm{v}}/N_{\mathrm{v}}^{\mathrm{suf}}$. We refer to such a diagram as a *relative sparsity-sampling* (RSS) diagram. Figure 4.4 shows RSS diagrams for strong and weak recovery, corresponding to the DT phase diagrams in Figure 4.3, created by reconstructing 100 **spikes** phantoms for $\kappa = 0.025$, $0.05, 0.1, 0.2, \ldots, 0.9$ and $N_{\mathrm{v}} = 2, 4, 6, \ldots, 32$.

Again, the variation of $N_{\mathrm{v}}^{\mathbf{L1}}$ over phantom instances is almost negligible, since there is a sharp transition from non-recovery to recovery. The relative sampling for recovery $\mu^{\mathbf{L1}}$ increases monotonically with the relative sparsity κ, although not in a linear way. As $\kappa \to 0$, the relative sampling for recovery $\mu^{\mathbf{L1}}$ also approaches 0, and similarly $\mu^{\mathbf{L1}} \to 1$ for $\kappa \to 1$, confirming that when the image is no longer sparse, **L1** gives no advantage over **L2**.

The RSS diagram also gives *quantitative* information on the recovery view number for **L1**. Assume, for example, that we are given an image of relative sparsity $\kappa = 0.1$, how many views would suffice for recovery? The RSS diagram shows that at $\kappa = 0.1$, we have $\mu^{\mathbf{L1}} = 0.31$, which corresponds to $N_{\mathrm{v}}^{\mathbf{L1}} = 8$ views. If the phantom has $\kappa = 0.6$, we obtain $\mu^{\mathbf{L1}} = 0.77$ and $N_{\mathrm{v}}^{\mathbf{L1}} = 20$ views. Note that the RSS-diagram works equally well for answering the opposite question, namely, what is the maximal
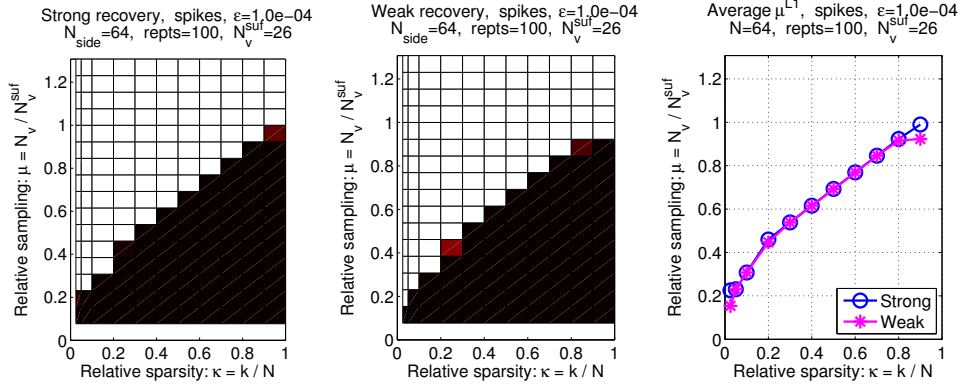
FIG. 4.4. *Left: Relative sparsity-sampling (RSS) diagram for strong recovery of the* **spikes** *class at $N_{\mathrm{side}} = 64$; the colorscale is as in Figure 4.3. Middle: Corresponding weak recovery RSS diagram. Right: Average relative sampling for strong and weak recovery over the phantom instances.*

relative sparsity that will allow recovery from 20 views?

To a great extent, the strong and weak recovery RSS diagrams in Figure 4.4 look similar, but small differences exist, e.g., at $\kappa = 0.2$. Recall that strong recovery can depend on the optimization algorithm used. To ensure strong recovery with MOSEK, we need $\mu = 0.46$ to obtain 100% recovery, while only 2% are strongly recovered at $\mu = 0.38$. For weak recovery, on the other hand, $\mu = 0.38$ suffices to recover 19% of the instances. The difference between the two types of recovery is even more pronounced at $\kappa = 0.9$, where only 13% have strong recovery at $\mu = 0.92$, while 100% are weakly recovered.

The similarities and differences between the RSS diagrams are emphasized in the rightmost plot in Figure 4.4, where the *average* $\mu^{\mathbf{L1}}$ over all instances at each κ is plotted. Since strong recovery of an instance implies weak recovery, the strong recovery curve will always be bounded from below by the weak recovery curve.

Clearly, the RSS diagram introduced here is strongly inspired by the DT phase diagram, but for several reasons we find the RSS diagram more intuitive to interpret for our CT applications:

1. A definition of the undersampling fraction relative to N, as used in the DT phase diagram, only makes sense when $M = N$ yields a full-rank matrix. This is not necessarily the case in CT, so a slightly different measure of undersampling, $\mu = N_{\mathrm{v}}/N_{\mathrm{v}}^{\mathrm{suf}}$, is required. (For the CT geometry studied here, $\delta \approx \mu$ because the N_{v} that yields full rank is close to $M = N$; but to be precise we still make the distinction.)

2. The sparsity fraction ρ for the DT phase diagram is relative to the number of measurements M, whereas our sparsity κ is relative to the number of pixels N. In the DT phase diagram, our relative sparsity is constant along hyperbolic curves instead of straight lines; to see this, note that a constant $\kappa = c$ is equivalent to $\rho\delta = c$, which is a hyperbola in the DT phase diagram. This means that the two diagrams are essentially different ways of visualizing the same data, only for slightly different ranges of the sparsity parameter. We find the RSS diagram more intuitive to use for our purposes, because the quantities of interest — the relative sparsity and sampling — can be read directly on the coordinate axes.

3. The DT phase diagram is typically used for studying randomly generated matrix instances A in addition to image instances x_{orig}, with the objective to understand recoverability over a whole matrix ensemble. For CT, on the other hand, we are interested in recoverability with a fixed matrix reflecting the given data acquisition geometry of the scanner in question, and the RSS diagram provides a mean for studying this situation: It attempts to answer how many views is sufficient for recovery of a phantom with a given (relative) sparsity for using a *fixed choice* of system matrix.

For these reasons, the remaining part of the article will solely use the RSS diagram to visualize the recovery results.
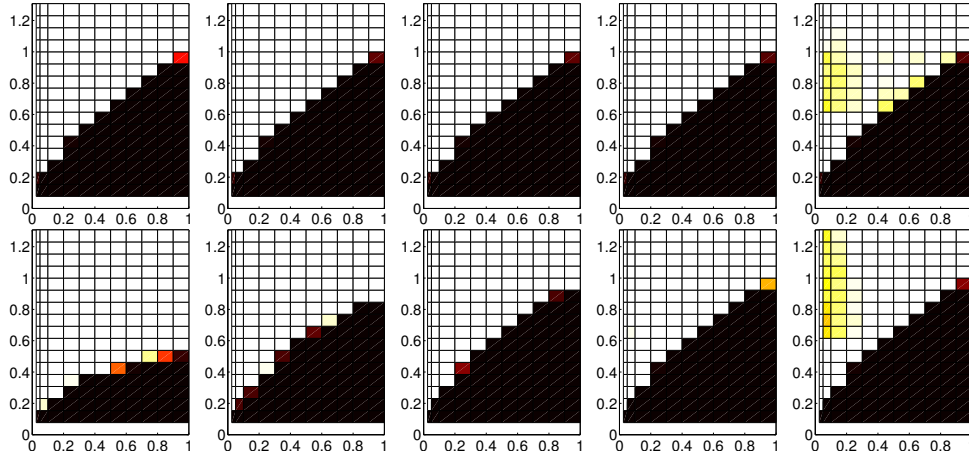

FIG. 4.5. *Strong (top) and weak (bottom) recovery RSS diagrams for class* **spikes**, $N_{\mathrm{side}} = 64$, *for* $\epsilon = 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}$ *(left to right). Axes labels and titles are omitted for improved visualization. Abscissas: relative sparsity* κ. *Ordinates: relative sampling* μ. *100 phantoms.*

4.4. Dependence on the threshold. The decision of recovery in the face of numerical computations (and rounding errors) depends on how close the computed solution is to the original, as measured by the threshold ϵ. As shown in Figure 4.2, the transition from non-recovery to recovery is often very distinct, and often recognized by a large jump in error norm (of five or more orders of magnitude)—although in some cases the transition is more gradual, cf. the lower left-most plot.

Figure 4.5 shows the strong and weak recovery RSS diagrams for a varying threshold $\epsilon = 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}$. Very little difference is seen among the three middle values; there are only minor variations on the location of the transition curve, due to few instances located on the border line. The threshold $\epsilon = 10^{-2}$ is too large, since the weak RSS diagram shows recovery for a too large number of instances. For the smallest threshold $\epsilon = 10^{-6}$, the transition curve remains similar to those in the mid-range; but certain instances well above the transition curve are not recovered. While this could indicate a real phenomenon, we are very confident that it arises when we approach the obtainable numerical accuracy of the optimization algorithm. We conclude that the appearance of the RSS diagram is not very sensitive to the choice of ϵ, and $\epsilon = 10^{-4}$ is a good choice for the chosen algorithm.

4.5. Dependence on image size. To study how the RSS diagram depends on image size, we construct additional diagrams for $N_{\mathrm{side}} = 32$ and 128. For $N_{\mathrm{side}} = 32$
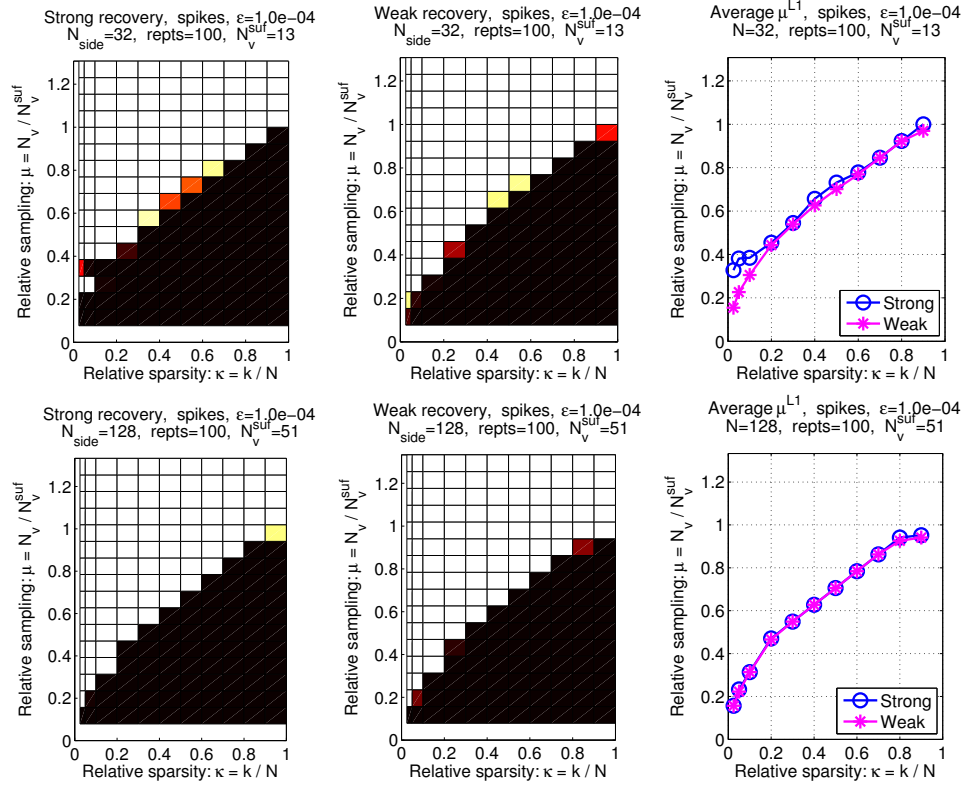
FIG. 4.6. *RSS diagram dependence with image size. Top: $N_{\text{side}} = 32$; bottom: $N_{\text{side}} = 128$; left: strong recovery; middle: weak recovery; right: average relative sampling for strong/weak recovery.*

we can use the same relative sampling as for $N_{\text{side}} = 64$ by taking $N_{\text{v}} = 1, 2, \ldots, 16$, since the matrix becomes full-rank at $N_{\text{v}}^{\text{suf}} = 13$. For $N_{\text{side}} = 128$ we have $N_{\text{v}}^{\text{suf}} = 51$, so by taking $N_{\text{v}} = 4, 8, \ldots, 64$ we obtain approximately the same relative sampling.

The two additional RSS diagrams are shown in Figure 4.6. Overall, we see the same monotone increase in $\mu^{\mathbf{L1}}$ with increasing κ. For $N_{\text{side}} = 32$, however, the transition from non-recovery to recovery is slightly more gradual, and the cases with the smallest κ have a larger $\mu^{\mathbf{L1}}$.

The following observation illustrates the need to distinguish between strong and weak recovery RSS diagrams. For $\kappa = 0.025$ and 0.05 the value $\mu = 0.23$ is sufficient for strong recovery, but adding one more view to obtain $\mu = 0.31$ destroys recovery. This seemingly counter-intuitive phenomenon is explained by the geometry underlying the data acquisition: Going from 3 to 4 views is not done by including an additional view to the existing views; rather the 4 views are distributed equi-angularly around the image, and hence the two system matrices are different. For this problem, 3 views provide enough data to recover the image, whereas the 4 views happen to be unfortunate. In the weak recovery RSS diagram, however, we see that both 3 and 4 views yield reliable weak recovery. We conclude that at 4 views the original image is indeed *a solution*, but it is non-unique and MOSEK happened to compute another solution than the one used to generate the data.

The RSS diagram for $N_{\text{side}} = 128$ is similar to the $N_{\text{side}} = 64$ case, except for

slightly better recovery at the extreme κ-values. This happens precisely where the strong and weak recovery RSS diagrams differ for $N_{\text{side}} = 64$, while at $N_{\text{side}} = 128$ we do not see this difference and accordingly the values of μ sufficient on average for strong and weak recovery agree. When comparing the three average-case figures, the weak recovery curves are almost identical, and the strong recovery curves approach the weak when N_{side} increases. The differences for $N_{\text{side}} = 32$ are most likely caused by discretization effects. Since the gap between the two curves closes, we conclude that for larger images the problem of having non-unique solutions is less pronounced. Moreover, $N_{\text{side}} = 64$ is sufficiently large to give representative results that can be extrapolated to larger N_{side}, especially when considering weak recovery.
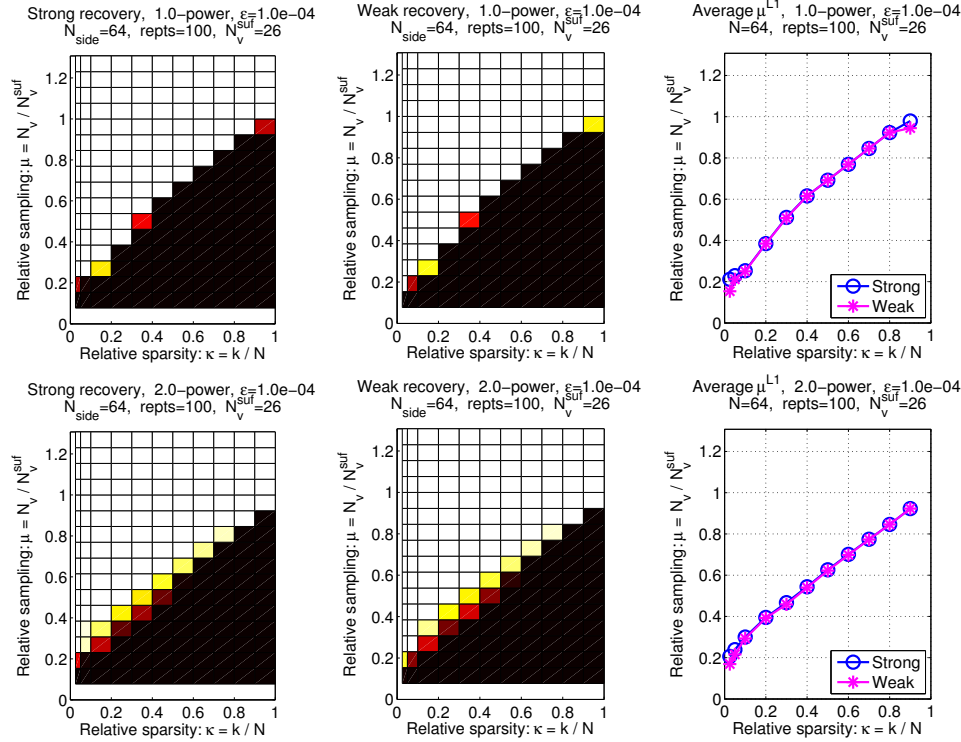

FIG. 4.7. *RSS diagrams for classes* **1.0-power** *(top) and* **2.0-power** *(bottom).*

4.6. Dependence on the phantom class. As argued in Section 2.2, we cannot expect recovery of *all* k-sparse images at a given relative sampling — at least only with *very few* non-zeros and unfavorably large relative sampling, due to the existence of pathological phantoms violating otherwise typical sufficient sampling. Hence, we study recoverability only for well-defined classes of phantom images.

Figure 4.7 shows RSS diagrams for the **1.0-power** and **2.0-power** classes for $N_{\text{side}} = 64$. Comparing with the **spikes** RSS diagrams in Figure 4.4 we observe similar trends. For **1.0-power** the transition from non-recovery to recovery occurs at the same (κ, μ)-values, but is slightly more gradual at a few κ-values. For **2.0-power** the transition is even more gradual, and occurs at lower μ-values for the mid and upper range of κ. We conclude that, on average, a smaller number of views suffices but the in-class recovery variability is larger. The **spikes** and **p-power** RSS diagrams

are similar, but there are also some differences and we conclude that sparsity does not fully determine recoverability; the structure of the nonzero locations also plays a role. The RSS diagrams can be used to study variation with structure and to determine if two classes have similar recoverability.

To further study how the recoverability depends on image class, we consider the RSS diagrams for the **signedspikes** class in Figure 4.8. Here, the transition is very sharp and the difference between strong and weak recovery is negligible. However, the transition occurs at much larger μ-values than for the **spikes** class in Figure 4.4. For example, at $\kappa = 0.4$ **spikes** has average $\mu^{\mathbf{L1}} = 0.62$ compared to 0.77 for **signedspikes**. At $\kappa = 0.8$ **spikes** still has undersampled recovery, although only at average $\mu^{\mathbf{L1}} = 0.92$, compared to no undersampling admitted for **signedspikes**. We conclude that **signedspikes** images are harder to recover, and the RSS diagram allows us to *quantify how much* harder, which we consider a useful property.
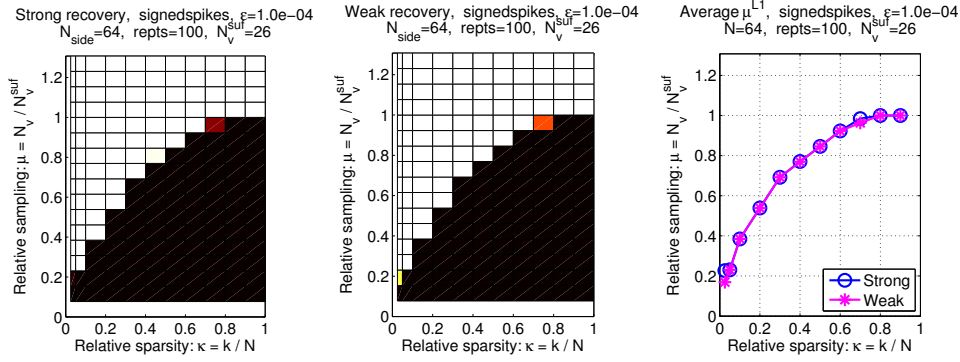

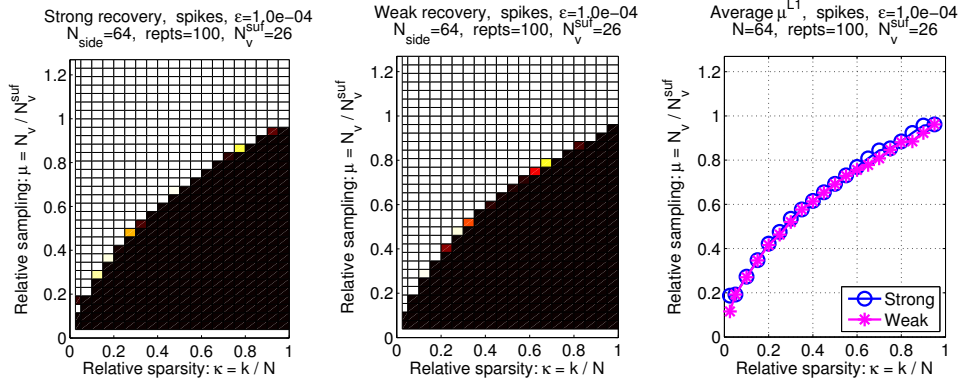FIG. 4.8. *RSS diagrams for the class* **signedspikes**.

FIG. 4.9. *High-resolution RSS diagrams for the* **spikes** *class at* $N_{\mathrm{side}} = 64$.

4.7. Dependence on diagram resolution. The RSS diagram can, among other things, be used to predict the recovery view number for images of a given relative sparsity. Assume the task is to recover a **spikes** image of expected relative sparsity 0.35; how many views should be used? The RSS diagrams in Figure 4.4 do not explicitly reveal the answer since $\kappa = 0.35$ is not included, but linear interpolation of the average $\mu^{\mathbf{L1}} = 0.54$ at $\kappa = 0.3$ and 0.62 at $\kappa = 0.4$ predicts $\mu^{\mathbf{L1}} = 0.58$.

We can verify this by increasing the resolution of the RSS diagram. Figure 4.9 shows diagrams for $\kappa = 0.025, 0.05, 0.10, 0.15, \ldots, 0.90, 0.95$ and $N_\mathrm{v} = 1, 2, 3, \ldots, 32$. Overall the trend agrees well with the lower-resolution RSS diagrams, but some finer details become apparent. Most notably, at $\kappa = 0.025$ we recognize the non-unique solution at 4 views that we saw for $N_\mathrm{side} = 32$ in Figure 4.6 (it seems that there is some intrinsic property of the 4-view geometry that causes the solution to be non-unique). Apart from that, we still observe a sharp transition from non-recovery to recovery throughout the κ-range. For $\kappa = 0.35$, we find the average relative sampling for recovery to be $\mu^{\mathbf{L1}} = 0.58$, thereby verifying our result interpolated from the lower-resolution RSS diagram.

In the interest of keeping the computational time required for determining the RSS diagrams reasonably low, we find — based on examples such as the given at $\kappa = 0.35$ — that the RSS diagram in Figure 4.4 is sufficiently well-resolved to locate the transition from non-recovery to recovery. If more computing resources are available, instead of increasing the diagram resolution it may be more advantageous to consider a larger ensemble of phantoms, to increase the reliability of the diagram.

4.8. Summary of results. We used simulations to demonstrate the use of the RSS diagram for empirical quantification of the relation between image sparsity and the number of views sufficient for **L1**-recovery from CT data. For the **spikes** class we observed a sharp transition from non-recovery to recovery throughout the relative sparsity range and, as function of relative sparsity, a monotonically increasing relative sampling for recovery. We saw that the RSS diagram is not sensitive to the threshold used in the recovery criterion, and that the transition from non-recovery to recovery is independent of the image size. We found similarities between RSS diagrams for different phantom classes, and the transition becomes smoother as the phantom structure increases. In summary, we believe that RSS diagrams are useful for understanding sparsity-exploiting reconstruction because they are:

- a structured framework for establishment and quantification of a relation between sparsity and sufficient sampling for a particular system,
- independent of any theoretical results guaranteeing solution uniqueness, and
- a way to study large systems through extrapolation from smaller systems.

In addition, the RSS diagram can easily be extended to studies of sparsity-exploiting methods with sparsity in other representations, with other penalty functions, relaxation of the equality constraint to admit noisy/inconsistent data, etc.

5. Discussion. Our CT simulation studies show that for a range of different sparse phantom classes it is possible to observe a sharp transition from non-recovery to recovery, both in the strong and weak sense. In light of the pessimistic theoretical recovery results mentioned in Section 2.2, it is perhaps surprising that such a sharp transition exists.

The simulation results speak only of the *average-case* recoverability, so there might still exist worst-case instances within our phantom classes that require far more samples than the average-case. Such pathological images are very unlikely to be realized in our simulations, so their influence can not be accounted for. On the other hand, such pathological phantoms might be very unrealistic from a practical perspective and therefore one can argue that they should be omitted for better understanding the behavior of the method when applied to "typical" images.

5.1. Limitations and extensions. The present ideal-case studies are far too unrealistic to provide any direct guidance on reconstruction quality and sampling in

a full-scale realistic CT application. Doing so is indeed the ultimate goal, but the problem has too many facets and too many parameters to analyze in enough detail to arrive at a complete understanding. Moreover, such results are necessarily very application-specific; for instance, the 2-norm metric (3.2) for image comparison is not appropriate for evaluating the practical utility of an image in a specific application.

Our intention here is to take a first step in this direction by proposing to carry out studies of particular systems of interest, and to provide an analysis for a simple but easily generalizable set-up. Our quantitative conclusions are, of course, only valid for the specific geometries, algorithms, data, and phantom classes.

The drawback of insisting on using a robust optimization algorithm is the limitation it enforces on the image size in the simulations; with MOSEK, we found $N_{\mathrm{side}} > 128$ to be impractical. We do not advocate MOSEK for larger systems than the ones considered here; faster algorithms that are applicable to large-scale problems exist, and potentially we only pay a price of reduced robustness. Also, the possibility to extrapolate RSS diagrams to larger image sizes as observed in Section 4.5 reduces the need for studies of larger systems. Since it appears that the relation between relative sparsity and relative sampling for recovery holds across the image size N_{side}, we do not need to study larger N_{side} but can simply extrapolate the relation to more realistic values of N_{side}.

5.2. Future work. The RSS diagram allows for generalization to increasingly more realistic set-ups. For example, other phantom classes can be considered with sparsity in other representations, such as in the gradient, and the penalty function can be changed to enforce the expected kind of sparsity. Noise and inconsistencies can be introduced in the data, the equality-constraint can be replaced by data misfit bounds controlled by regularization parameters, and the system can be changed (e.g., to a limited angle CT problem). Making such generalizations might require modifications of the sparsity measure and the recovery criterion.

Earlier studies of TV reconstructions [18] seem to show a relation between gradient image sparsity and sufficient sampling, but due to the complexity of the test problems we found it difficult to understand and interpret the influence of each experiment. An investigation based on the RSS diagram could provide more structured insight. For instance, we might learn that TV reconstructions of a class of "blocky" phantoms exhibit a well-defined recovery-curve similar to the one in the present study.

We always face the problem of possible non-unique solutions to **L1**, leading to RSS and DT phase diagrams that depend on the particular choice of optimization algorithm. To that end, we introduced the weak recovery RSS diagram which is independent of the optimization algorithm, and we found that in most cases the weak recovery curves provide tight lower bounds on the strong recovery curves. We are therefore fairly confident that the MOSEK results provide an accurate picture of the recoverability. It would be interesting to also determine a transition curve for the existence of a *unique* solution, since such a curve provides an upper bound the strong recovery curve. With a unique solution, the choice of optimization algorithm is not important as long as a robust optimization algorithm is used. We expect that **L1**-uniqueness can be investigated by numerically verifying a set of necessary and sufficient conditions [15]. We did not pursue that idea in the present work in order to focus on an empirical approach easily generalizable to other penalties, such as TV, for which similar conditions are not available.

Another interesting future direction is to study the in-class recovery variability, i.e., why the **2.0-power** class transition from non-recovery to recovery is more grad-

ual. Would it be possible to identify differences between instances that were recovered and ones that were not, e.g., in the spatial location of the non-zero pixels, or in the histograms of pixel values? This could lead to subdividing the phantom class into partitions each having sharper transitions occurring at different relative sampling values and thereby an even better understanding of what factors influence the recoverability.

6. Conclusion. Inspired by the Donoho-Tanner phase diagram we devised a relative sparsity-sampling diagram for studying empirical recoverability in sparsity-exploiting X-ray CT image reconstruction. Our approach is not limited to sparsity in a specific domain or reconstruction by solving a specific optimization problem. We focused on sparsity in the pixel domain and reconstruction using the ℓ_1-norm penalty.

Our numerical simulations demonstrate a pronounced relation between image sparsity and the number of projections needed for recovery, for a range of image classes without and with structure and classes with signed and un-signed pixel values. In the majority of the studied cases, we found a sharp transition from non-recovery to recovery — a result that hitherto, to our knowledge, has not been established for CT. The sharp transition allows for quantitatively predicting the number of projections that, on average, suffices for accurate ℓ_1-recovery of phantom images from a specific class, or conversely, to determine the maximal sparsity of an image that can be recovered for a certain number of views. With these initial results we have taken a step towards better quantitative understanding of the recoverability from undersampled measurements in X-ray CT, and additionally we provide a tool for determining similar answers for increasingly realistic systems.

REFERENCES

[1] L. APPLEBAUM, S. D. HOWARD, S. SEARLE, AND R. CALDERBANK, *Chirp sensing codes: Deterministic compressed sensing measurements for fast recovery*, Appl. Comput. Harmon. Anal., 26 (2009), pp. 283–290.

[2] H. H. BARRETT AND K. J. MYERS, *Foundations of Image Science*, John Wiley & Sons, Hoboken, NJ, 2004.

[3] M. BEISTER, D. KOLDITZ, AND W. A. KALENDER, *Iterative reconstruction methods in X-ray CT*, Physica Med., 28 (2012), pp. 94–108.

[4] J. BIAN, J. H. SIEWERDSEN, X. HAN, E. Y. SIDKY, J. L. PRINCE, C. A. PELIZZARI, AND X. PAN, *Evaluation of sparse-view reconstruction from flat-panel-detector cone-beam CT*, Phys. Med. Biol., 55 (2010), pp. 6575–6599.

[5] A. M. BRUCKSTEIN, D. L. DONOHO, AND M. ELAD, *From sparse solutions of systems of equations to sparse modeling of signals and images*, SIAM Rev., 51 (2009), pp. 34–81.

[6] E. J. CANDÈS, J. ROMBERG, AND T. TAO, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*, IEEE Trans. Inform. Theory, 52 (2006), pp. 489–509.

[7] E. J. CANDÈS AND T. TAO, *Decoding by linear programming*, IEEE Trans. Inform. Theory, 51 (2005), pp. 4203–4215.

[8] K. CHOI, J. WANG, L. ZHU, T.-S. SUH, S. BOYD, AND L. XING, *Compressed sensing based cone-beam computed tomography reconstruction with a first-order method*, Med. Phys., 37 (2010), pp. 5113–5125.

[9] D. DONOHO AND J. TANNER, *Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing.*, Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci., 367 (2009), pp. 4273–4293.

[10] D. L. DONOHO, *Compressed sensing*, IEEE Trans. Inform. Theory, 52 (2006), pp. 1289–1306.

[11] D. L. DONOHO AND M. ELAD, *Optimally sparse representation in general (non-orthogonal) dictionaries via L1 minimization*, Proc. Natl. Acad. Sci. USA, 100 (2003), pp. 2197–2202.

[12] C. DOSSAL, G. PEYRÉ, AND J. FADILI, *A numerical exploration of compressed sampling recovery*, Linear Algebra Appl., 432 (2010), pp. 1663–1679.

[13] M. ELAD, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, Springer, New York, NY, 2010.

[14] L. A. FELDKAMP, L. C. DAVIS, AND J. W. KRESS, *Practical cone-beam algorithm*, J. Opt. Soc. Amer. A, 1 (1984), pp. 612–619.

[15] M. GRASMAIR, M. HALTMEIER, AND O. SCHERZER, *Necessary and sufficient conditions for linear convergence of L1-regularization*, Comm. Pure Appl. Math., 64 (2011), pp. 161–182.

[16] X. HAN, J. BIAN, D. R. EAKER, T. L. KLINE, E. Y. SIDKY, E. L. RITMAN, AND X. PAN, *Algorithm-enabled low-dose micro-CT imaging*, IEEE Trans. Med. Imaging, 30 (2011), pp. 606–620.

[17] P. C. HANSEN AND M. SAXILD-HANSEN, *AIR Tools – A MATLAB package of algebraic iterative reconstruction methods*, J. Comput. Appl. Math., 236 (2012), pp. 2167–2178.

[18] J. H. JØRGENSEN, E. Y. SIDKY, AND X. PAN, *Quantifying admissible undersampling for sparsity-exploiting iterative image reconstruction in X-ray CT*, Arxiv preprint arXiv:1109.0629, (2012).

[19] M. LI, H. YANG, AND H. KUDO, *An accurate iterative reconstruction algorithm for sparse objects: application to 3D blood vessel reconstruction from a limited number of projections*, Phys. Med. Biol., 47 (2002), pp. 2599–2609.

[20] MOSEK APS, *MOSEK Optimization Software (www.mosek.com)*.

[21] F. NATTERER, *The Mathematics of Computerized Tomography*, John Wiley & Sons, New York, NY, 1986.

[22] X. PAN, E. Y. SIDKY, AND M. VANNIER, *Why do commercial CT scanners still employ traditional, filtered back-projection for image reconstruction?*, Inverse Problems, 25 (2009), p. 123009.

[23] S. PETRA AND C. SCHNÖRR, *TomoPIV meets compressed sensing*, Pure Math. Appl. (PU.M.A.), 20 (2009), pp. 1737–1739.

[24] M. E. PFETSCH AND A. M. TILLMANN, *The computational complexity of the restricted isometry property, the nullspace property, and related concepts in compressed sensing*, Arxiv preprint arXiv:1205.2081, (2012).

[25] N. PUSTELNIK, C. DOSSAL, F. TURCU, Y. BERTHOUMIEU, AND P. RICOUX, *A greedy algorithm to extract sparsity degree for L1/L0-equivalence in a deterministic context*, in Proceedings of EUSIPCO, Bucharest, Romania, 2012.

[26] I. REISER AND R. M. NISHIKAWA, *Task-based assessment of breast tomosynthesis: Effect of acquisition parameters and quantum noise*, Med. Phys., 37 (2010), pp. 1591–1600.

[27] L. RITSCHL, F. BERGNER, C. FLEISCHMANN, AND M. KACHELRIESS, *Improved total variation-based CT image reconstruction applied to clinical data*, Phys. Med. Biol., 56 (2011), pp. 1545–1561.

[28] L. I. RUDIN, S. OSHER, AND E. FATEMI, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268.

[29] E. Y. SIDKY, C.-M. KAO, AND X. PAN, *Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT*, J. X-Ray Sci. Technol., 14 (2006), pp. 119–139.

[30] E. Y. SIDKY AND X. PAN, *Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization*, Phys. Med. Biol., 53 (2008), pp. 4777–4807.

[31] L. YU, X. LIU, S. LENG, J. M. KOFLER, J. C. RAMIREZ-GIRALDO, M. QU, J. CHRISTNER, J. G. FLETCHER, AND C. H. MCCOLLOUGH, *Radiation dose reduction in computed tomography: techniques and future perspective.*, Imaging Med., 1 (2009), pp. 65–84.